DroMOOC

Computer Vision

Basic Level

Image Features

Antoine Manzanera

ENSTA

Image Features

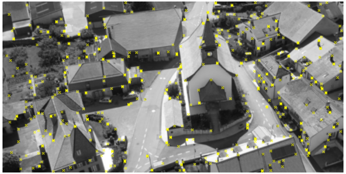
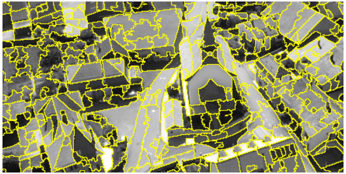
Objectives of the Lecture

Know the basics of feature extraction and image representation for computer vision: why and how to reduce the information support, and how to use it for different purposes.

Outline of the Lecture

- Image features and Local geometry
- Scale-space derivatives
- Ex.1: Contours
- Ex.2: Corner points
- Ex.3: Blob points
- Feature description

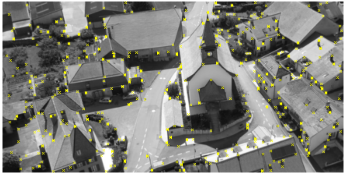
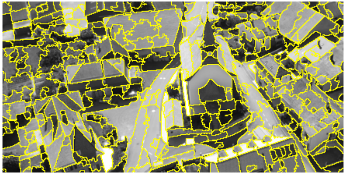
Image Features: Why and What?

The purpose is to efficiently represent image parts, so that they can be matched between their occurrences within video sequences or stereo pairs. It consists in:

- Reducing the information support to a compact and significant subset.
- Representing this subset by a robust, discriminant and small descriptor.

Feature detection

Feature detection is the process that extracts the relevant pixels in images.
In the above examples: keypoints, regions, contour points. A good detector should be:

- Representative (Provide many points...)
- Repeatable (...stable under deformation)
- Fast!

Feature Description

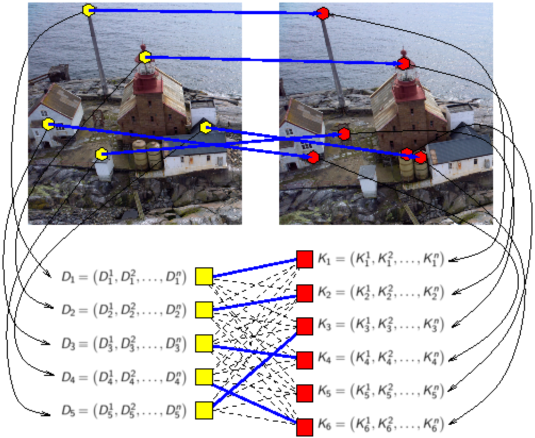
Feature description is the process that attaches to each selected pixel a representation vector called *descriptor* so that matching features can be done by comparing descriptors (e.g. using a distance measure). A good descriptor should be:

- *Compact* (Low dimension) for efficiency purpose.
- *Distinctive* for discrimination purpose.
- *Invariant* to image change (e.g. pose and illumination)

Examples: neighbourhood patches, Contrast, Orientation, Curvature, Histograms of Orientation...

Feature points matching

Many computer vision applications are based on matching a collection of feature points, which is usually performed using nearest neighbour search and consistency criteria.



Typical use cases:

- Sparse optical flow
- Object tracking
- Image registration
- Visual servoing
- Visual odometry
- Object recognition

Image Features and Local Geometry

Image features are related to spatially *salient* structures, characterised by their local geometry, which is well described using spatial derivatives.

For example, at order 1, the gradient vector $\nabla I = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}\right)^T$ provides two essential features:

- Orientation: $\arg \nabla I = \arctan \left(\frac{\partial I/\partial y}{\partial I/\partial x}\right)$
- Contrast: $||\nabla I|| = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2}$

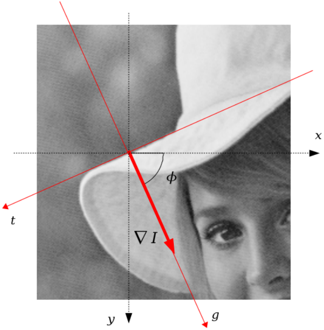
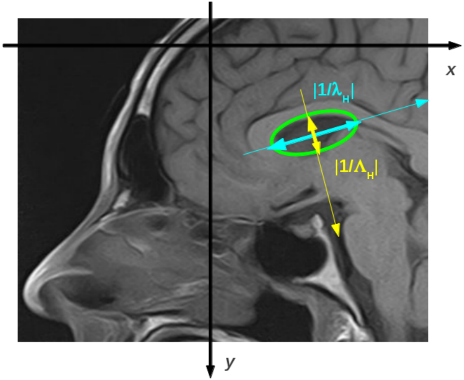


Image Features and Local Geometry

At order 2, the Hessian matrix $H_I = \begin{pmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial y^2} \end{pmatrix}$ provides through its eigen vectors (resp.values) the main directions (resp. intensities) of curvature.



Scale-space derivatives

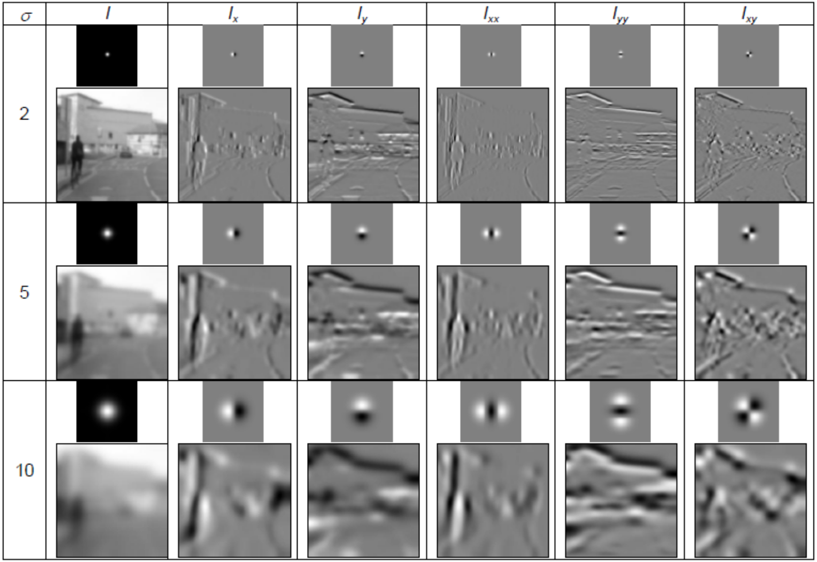
- First lecture: finite difference convolution based derivatives.
- Derivability $\Leftrightarrow$ Regularity.
- Scale space principle: on discrete images, the derivative only makes sense up to a scale parameter.
- The level of regularity is explicitly enforced by Gaussian smoothing.

Gaussian derivatives

$$\left(\frac{\partial^{i+j} I}{\partial^i x \partial^j y}\right)_\sigma \equiv I \star \frac{\partial^{i+j} G_\sigma}{\partial^i x \partial^j y}$$

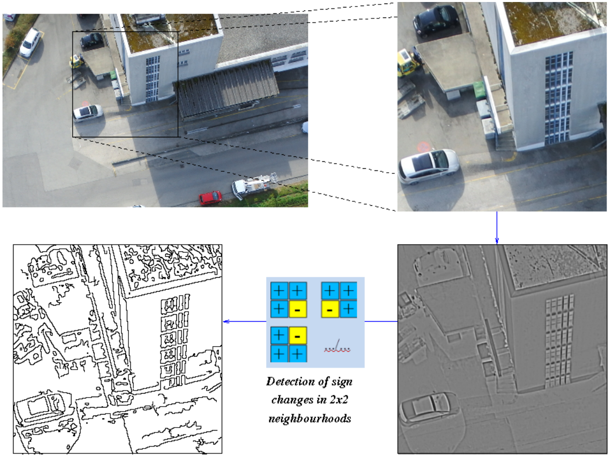
with $G_\sigma = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$.

Scale-space derivatives



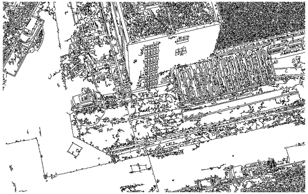
Example 1: Contour detection

- Contour: location of significant changes in the image.
- Derivative formulation: Zero-crossings of the Laplacian $\Delta I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$.

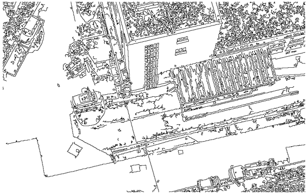


Zero−crossings   *Laplacian* $(\sigma = 2.0)$

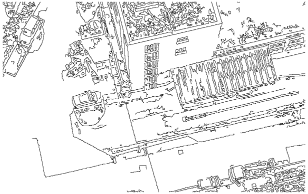
Multi-scale Contours



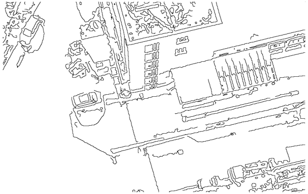
$\sigma = 1.0$



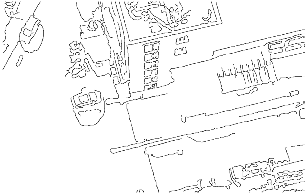
$\sigma = 1.5$



$\sigma = 2.0$



$\sigma = 2.0$



$\sigma = 4.0$

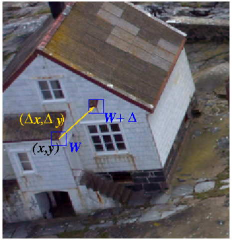
$\sigma = 6.0$

Example 2: Corner point detection

- A corner point is a locus where the value varies significantly in different directions.
- It can be evaluated at any location (x, y), using the local dissimilarity measure $\chi(x, y)$:

$$\chi(x, y) = \sum_{(i,j)\in W} (I(i, j) - I(i + \Delta x, j + \Delta y))^2$$

- where $(\Delta x, \Delta y)$ is a 2d displacement vector.
- where W is a small neighbourhood window centred on pixel (x, y).



Example 2: Corner point detection

Using a first order approximation of $I(i+\Delta x, j+\Delta y)$, we get:

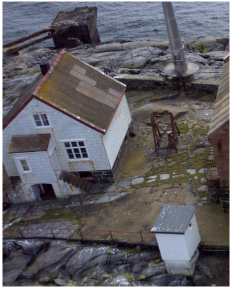
$$\chi(x,y) \simeq \begin{pmatrix}\Delta x & \Delta y\end{pmatrix} \underbrace{\begin{pmatrix} \displaystyle\sum_{(i,j)\in W} \left(\frac{\partial I}{\partial x}(i,j)\right)^2 & \displaystyle\sum_{(i,j)\in W} \left(\frac{\partial I}{\partial x}(i,j).\frac{\partial I}{\partial y}(i,j)\right) \\ \displaystyle\sum_{(i,j)\in W} \left(\frac{\partial I}{\partial x}(i,j).\frac{\partial I}{\partial y}(i,j)\right) & \displaystyle\sum_{(i,j)\in W} \left(\frac{\partial I}{\partial y}(i,j)\right)^2 \end{pmatrix}}_{\Xi(x,y)} \begin{pmatrix}\Delta x \\ \Delta y\end{pmatrix}$$

- $\Xi(x,y)$ is called the structure matrix at pixel (x,y).
- The corner points are those for which $\Xi(x,y)$ has *two large* eigen values.

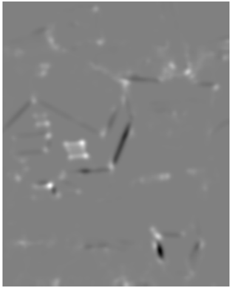
Example 2: Harris corner point detector

The Harris detector does not explicitly calculate the eigen values of Ξ, but uses the interest function Θ instead:

$$\Theta(x, y) = \det \Xi(x, y) - \alpha \mathrm{trace} \Xi(x, y)$$



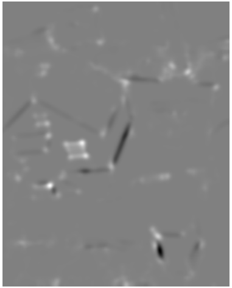
Original I



Interest Function Θ ($\sigma = 3.0$)

Example 2: Harris corner point detector

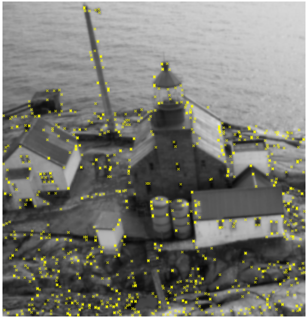
- The local maxima of function Θ are the corner points.
- The intrinsic scale of the corner points is the one used to estimate the partial derivatives.



Interest Function Θ ($\sigma = 3.0$)

Harris corner points

Multi-scale Harris Corner Point Detector 1/2



$\sigma = 1.0$

$\sigma = 2.0$

$\sigma = 3.0$

Multi-scale Harris Corner Point Detector 2/2

$\sigma = 4.0$

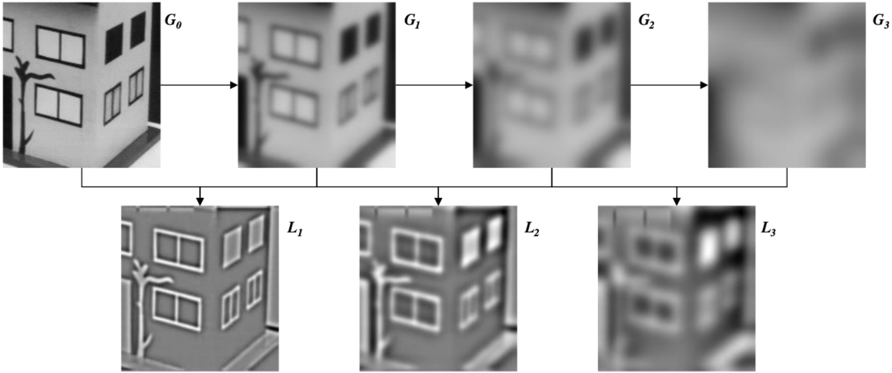
$\sigma = 5.0$

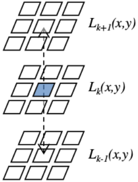
$\sigma = 6.0$

Example 3: Blob detectors

- *Blobs* are elliptical bright or dark structures in the image.
- They are usually better suited than corner points to *large scales*.
- Unlike corner points, they are characterised at *second order* of derivation.
- They can be detected as local maxima of the *Hessian determinant* (e.g. SURF points).
- They can also be detected as the local maxima of the *multiscale Difference-of-Gaussians* (e.g. SIFT points)

SIFT Blob detector





The SIFT points are the local extrema in the 3d scale-space (x, y, k) of the normalised $L_k(x, y)$ Difference of Gaussians function.

SIFT Blob detector

- Each SIFT keypoint is localised in both space (x, y) and scale k.
- It has its specific orientation: $\theta(x, y, k) = \arctan \frac{\frac{\partial}{\partial y} G_k(x,y)}{\frac{\partial}{\partial x} G_k(x,y)}$.

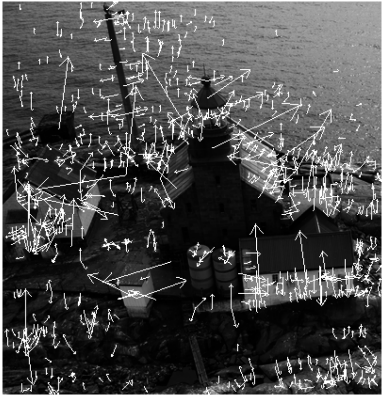


Image 1: 984 keypoints

Image 2: 792 keypoints

SURF Blob detector

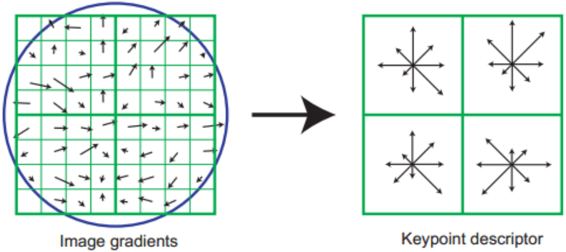
SURF is a fast alternative to blob detection. It calculates the multi-scale determinant of the Hessian, by approximating the second derivatives by fast convolutions using rectangular (Haar-like) kernels:

$$\underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -2 & -2 & -2 & -2 & -2 & -2 & -2 \\ -2 & -2 & -2 & -\mathbf{2} & -2 & -2 & -2 \\ -2 & -2 & -2 & -2 & -2 & -2 & -2 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}}_{H_{yy}} \quad \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & \mathbf{1} & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \end{pmatrix}}_{H_{xy}}$$

- $\frac{\partial^2 I}{\partial y^2} \simeq I \star H_{yy}$
- $\frac{\partial^2 I}{\partial x \partial y} \simeq I \star H_{xy}$
- etc.

SIFT Feature Descriptor

- The SIFT feature descriptor is a collection of Gradient Orientation Histograms collected over square regions around each SIFT point.
- For rotation invariance purpose, the specific orientation of the SIFT point is used as origin of all histograms.



SIFT Feature Matching

- The matching between Image 1 (984 keypoints) and Image 2 (792 keypoints) provides 151 matches considered as "correct".
- A match is considered correct here if the ratio between the distance to the closest descriptor and the distance to the second closest descriptor is less than a certain threshold (e.g. 0.7).

Image Features - CONCLUSION

What have we learned so far

- Interest of image features
- Scale-space feature detection
- Feature description

Upcoming Lecture

- Feature matching
- Video: Optical Flow
- Spatial regularisation